空間自相關：類別變數
Spatial Autocorrelation for categorical variables

授課教師：溫在弘
E-mail: wenthung@ntu.edu.tw

複習：Spatial Autocorrelation: Moran’s I Statistic

Product of the deviation from the mean for all pairs of adjacent regions ($w_{ij}=1$)

$$I = \frac{n\sum_{i=1}^{n}\sum_{i=j}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\left(\sum_{i=1}^{n}(y_i - \bar{y})^2\right)\left(\sum\sum_{i\neq j} w_{ij}\right)}$$

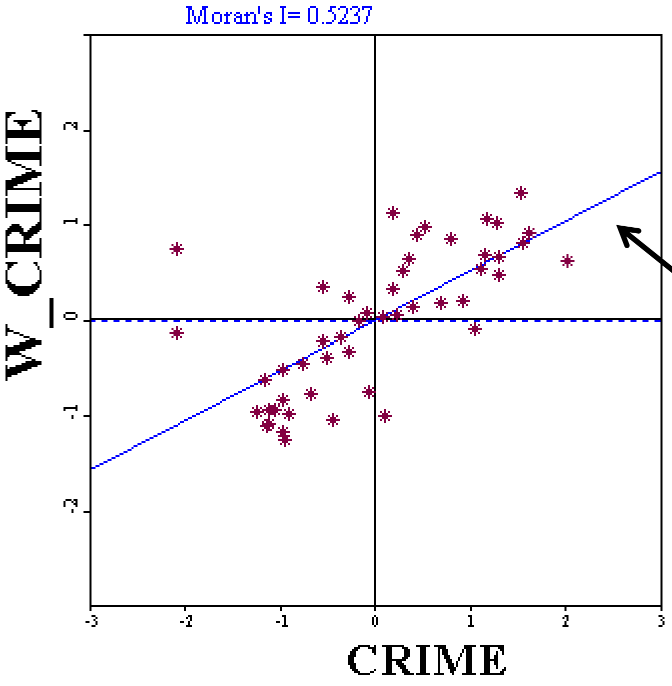
Essentially a measure of variance across the regions

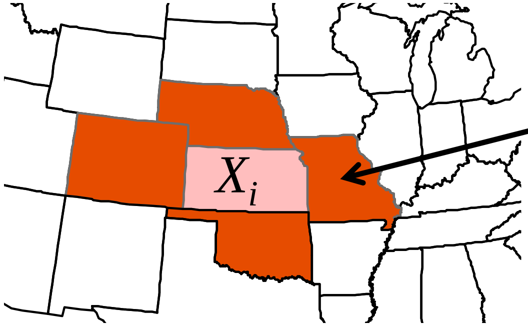
Sum of the weights (count of all adjacent pairs)

- n = number of regions
- w_{ij} = measure of spatial proximity between region i and j

複習：Moran Scatter Plots

Moran’s I can be interpreted as the correlation between variable, X, and the “spatial lag” of X formed by averaging all the values of X for the neighboring polygons.





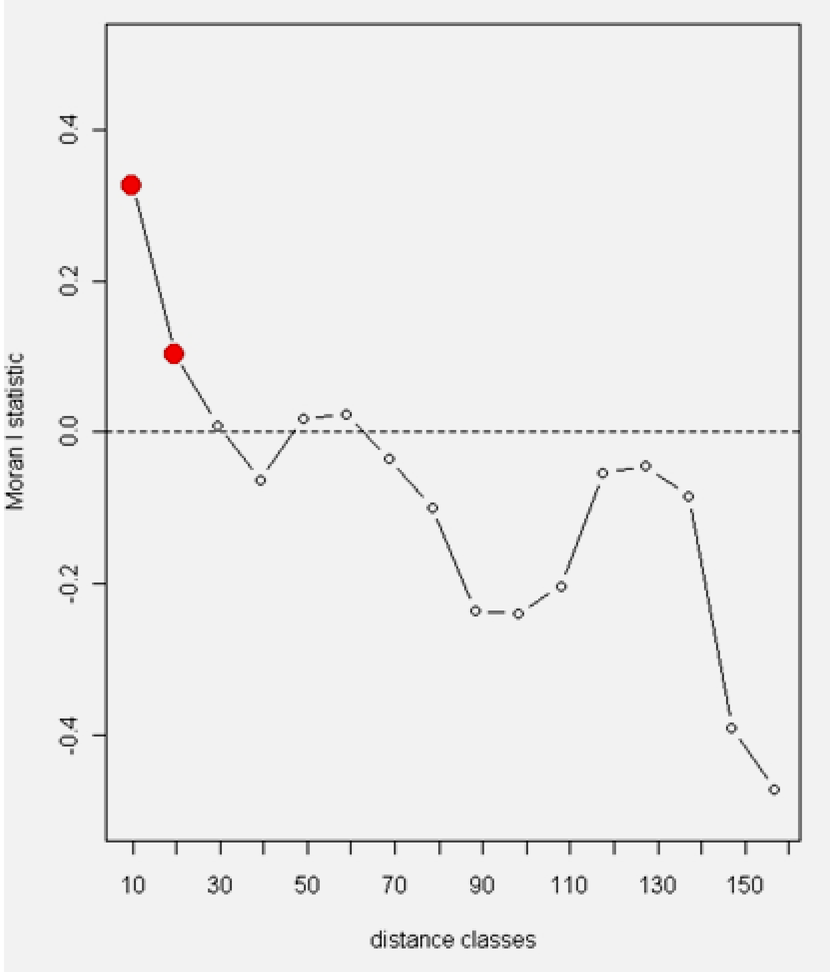
Lag X_i is average of these

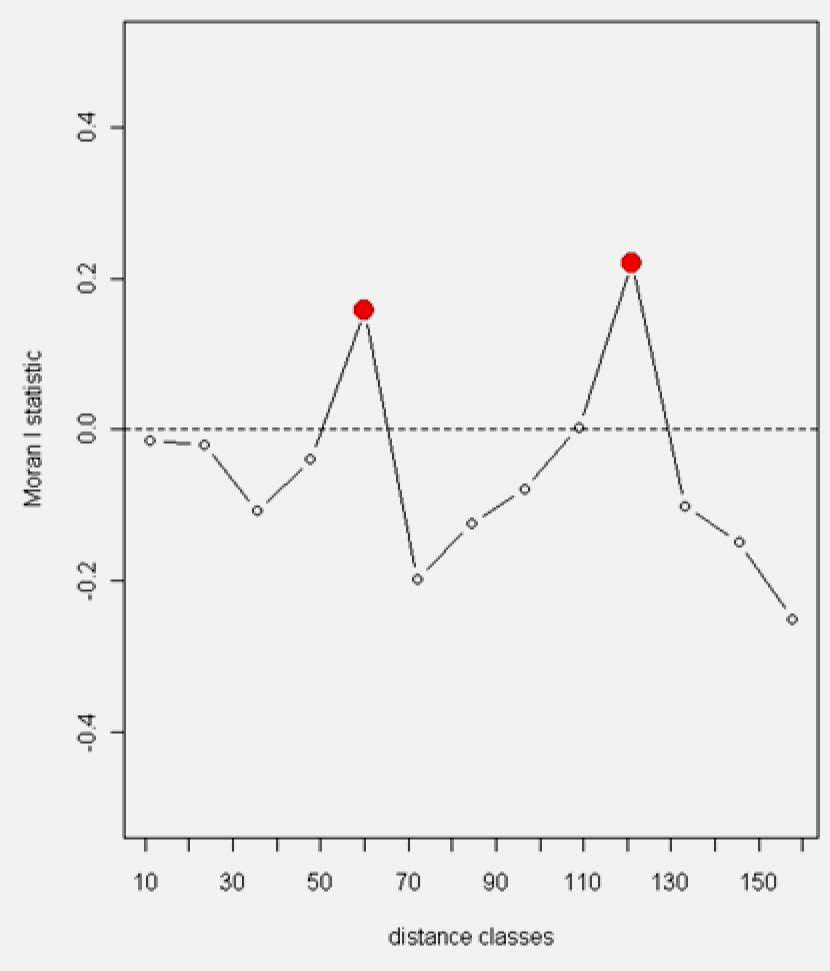
Least squares “best fit” line to the points.

The slope of this *regression line* is Moran’s I

複習： Moran Correlograms

Correlogram: plot distance on X-axis against correlation coefficient on Y-axis





複習：Getis-Ord General G-statistic

- Moran’s I 無法區別

  “ hot spots” or “cold spots”

- ***Spatial Concentration*** method
- Definition

$$G(d) = \frac{\sum\sum w_{ij}(d)x_i x_j}{\sum\sum x_i x_j}$$

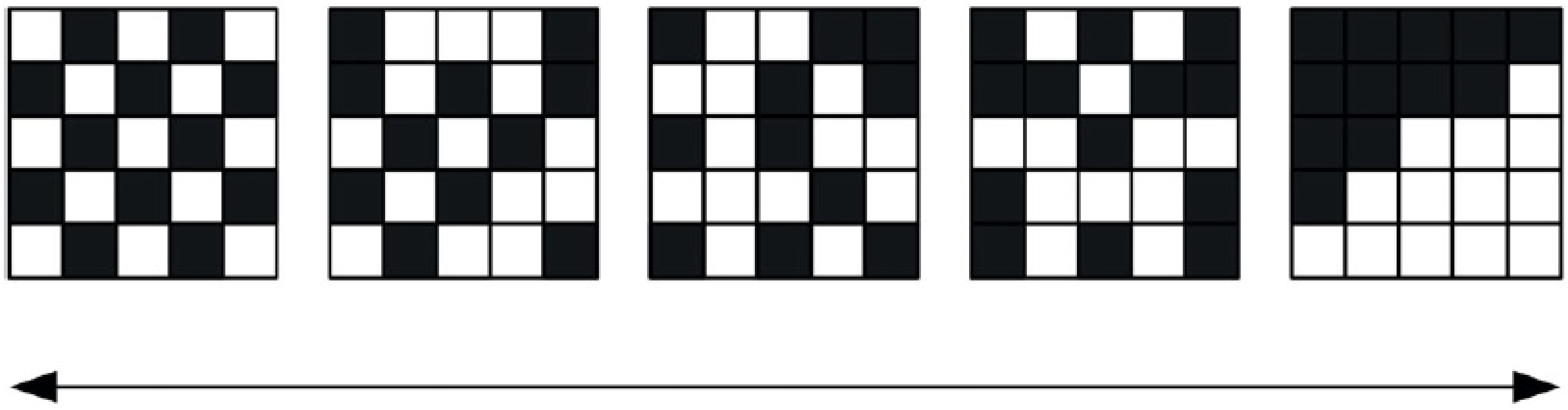
d : neighborhood distance

W_{ij} : 1 if it is within d, 0 otherwise

- Calculation of G must begin by identifying a neighborhood distance within which cluster is expected to occur

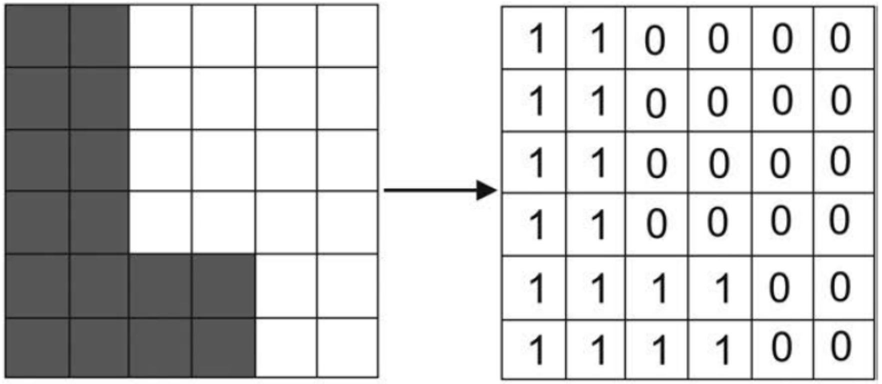
Join (Joint) Counts Statistics

- It is used for binary nominal data such as 1/0, yes/no and urban/rural counties.
- The binary variable is denoted by two colors, black (B) and white (W)
- The method measures the spatial relationships between similar or dissimilar attributes in adjacent areas.



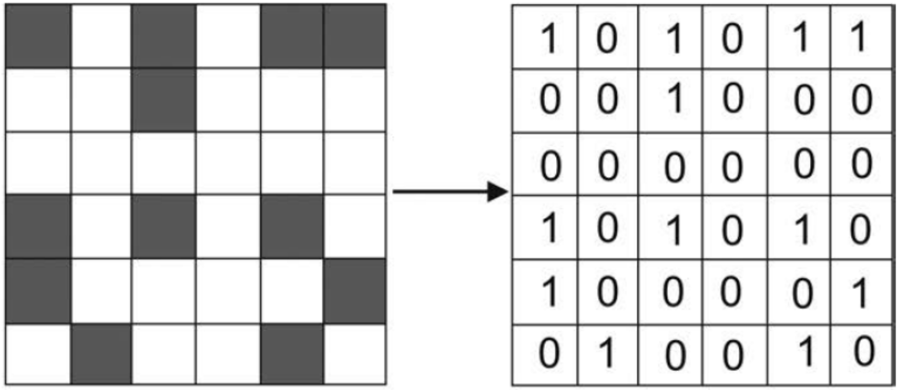
Join (Joint) Counts Statistics

- If a given attribute of 1 occurs in an area, then the area will be assigned B. If it does not and has an attribute of 0, then it will be assigned W.
- If two neighboring areas share a common boundary, they are conceptualized as joined.



1	1	0	0	0	0
1	1	0	0	0	0
1	1	0	0	0	0
1	1	0	0	0	0
1	1	1	1	0	0
1	1	1	1	0	0

Areal Pattern A



1	0	1	0	1	1
0	0	1	0	0	0
0	0	0	0	0	0
1	0	1	0	1	0
1	0	0	0	0	1
0	1	0	0	1	0

Areal Pattern B

Join counts

BB (black–black) joins:

$$BB = \frac{1}{2}\sum_i \sum_j w_{ij} x_i x_j$$

BW (black–white) joins:

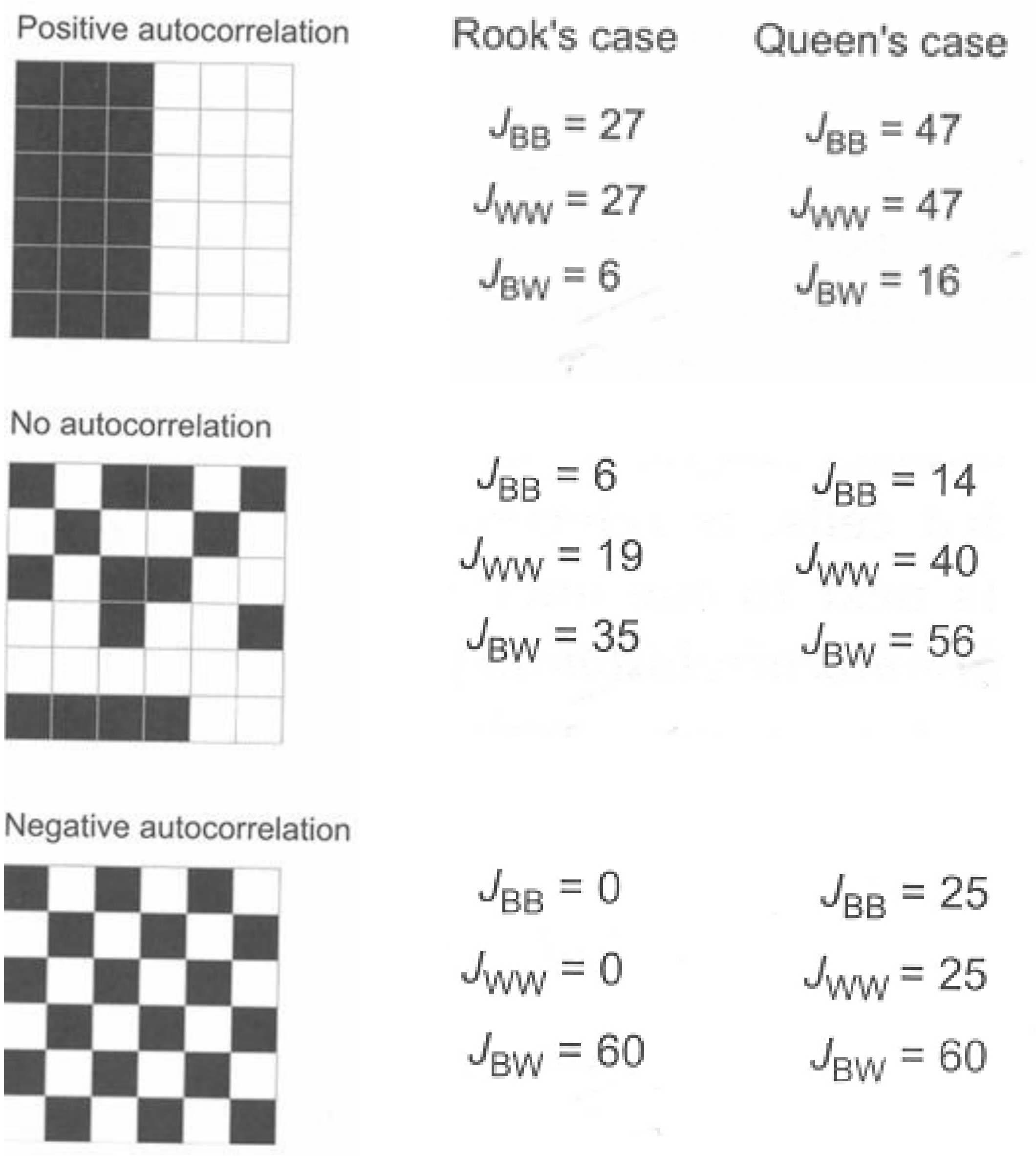
$$BW = \frac{1}{2}\sum_i \sum_j w_{ij} (x_i - x_j)^2$$

WW (white–white) joins:

$$WW = \frac{1}{2}\sum_i \sum_j w_{ij} (1 - x_i)(1 - x_j)$$

where x_i is the observer value for variant X_i, $x_i = 1$ when the ith area is B, and $x_i = 0$ when the ith area is W, and w_{ij} is weight for each pair of objects i and j.

Examples



	Rook's case	Queen's case
Positive autocorrelation	$J_{BB} = 27$	$J_{BB} = 47$
	$J_{WW} = 27$	$J_{WW} = 47$
	$J_{BW} = 6$	$J_{BW} = 16$
No autocorrelation	$J_{BB} = 6$	$J_{BB} = 14$
	$J_{WW} = 19$	$J_{WW} = 40$
	$J_{BW} = 35$	$J_{BW} = 56$
Negative autocorrelation	$J_{BB} = 0$	$J_{BB} = 25$
	$J_{WW} = 0$	$J_{WW} = 25$
	$J_{BW} = 60$	$J_{BW} = 60$

Large proportion of BB and WW joins

Different numbers of BW, BB and WW joins

Large number of BW joins;
Small number of BB and WW joins

Join Count: Test Statistic

$$Z = \frac{\text{Observed} - \text{Expected}}{\sigma_{\text{Expected}}}$$

- Each of the null hypotheses for the three types of joins determines whether the compared differences are statistically significant at p-value less than 0.05.
- This is done by calculating the Z-test for each join and deciding whether the null hypothesis is true.

Join Count: Test Statistic

$$Z = \frac{\text{Observed} - \text{Expected}}{\sigma_{\text{Expected}}}$$

Expected = random pattern generated by tossing a coin in each cell.

Expected Counts

$$E(J_{\mathrm{BB}}) = kp_{\mathrm{B}}^2$$

$$E(J_{\mathrm{WW}}) = kp_{\mathrm{W}}^2$$

$$E(J_{\mathrm{BW}}) = 2kp_B p_W$$

Based on free (normality) sampling

Standard Deviation of Expected Counts (Standard Error)

$$E(s_{\mathrm{BB}}) = \sqrt{kp_{\mathrm{B}}^2 + 2mp_{\mathrm{B}}^3 - (k+2m)p_{\mathrm{B}}^4}$$

$$E(s_{\mathrm{WW}}) = \sqrt{kp_{\mathrm{W}}^2 + 2mp_{\mathrm{W}}^3 - (k+2m)p_{\mathrm{W}}^4}$$

$$E(s_{\mathrm{BW}}) = \sqrt{2(k+m)p_{\mathrm{B}}p_{\mathrm{W}} - 4(k+2m)p_{\mathrm{B}}^2 p_{\mathrm{W}}^2}$$

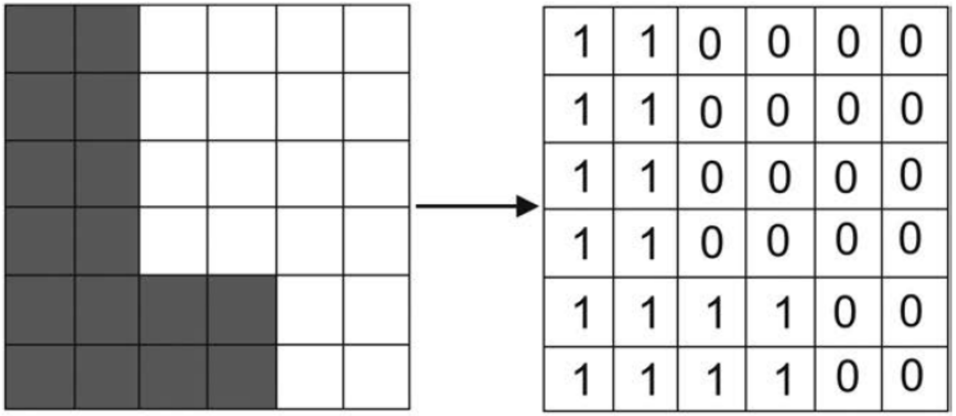
where: k is the total number of joins (neighbors)

p_B is the expected proportion Black, if random

p_W is the expected proportion White

m is calculated from k according to: $m = \frac{1}{2}\sum_{i=1}^{n} k_i(k_i - 1)$

Example



1	1	0	0	0	0
1	1	0	0	0	0
1	1	0	0	0	0
1	1	0	0	0	0
1	1	1	1	0	0
1	1	1	1	0	0

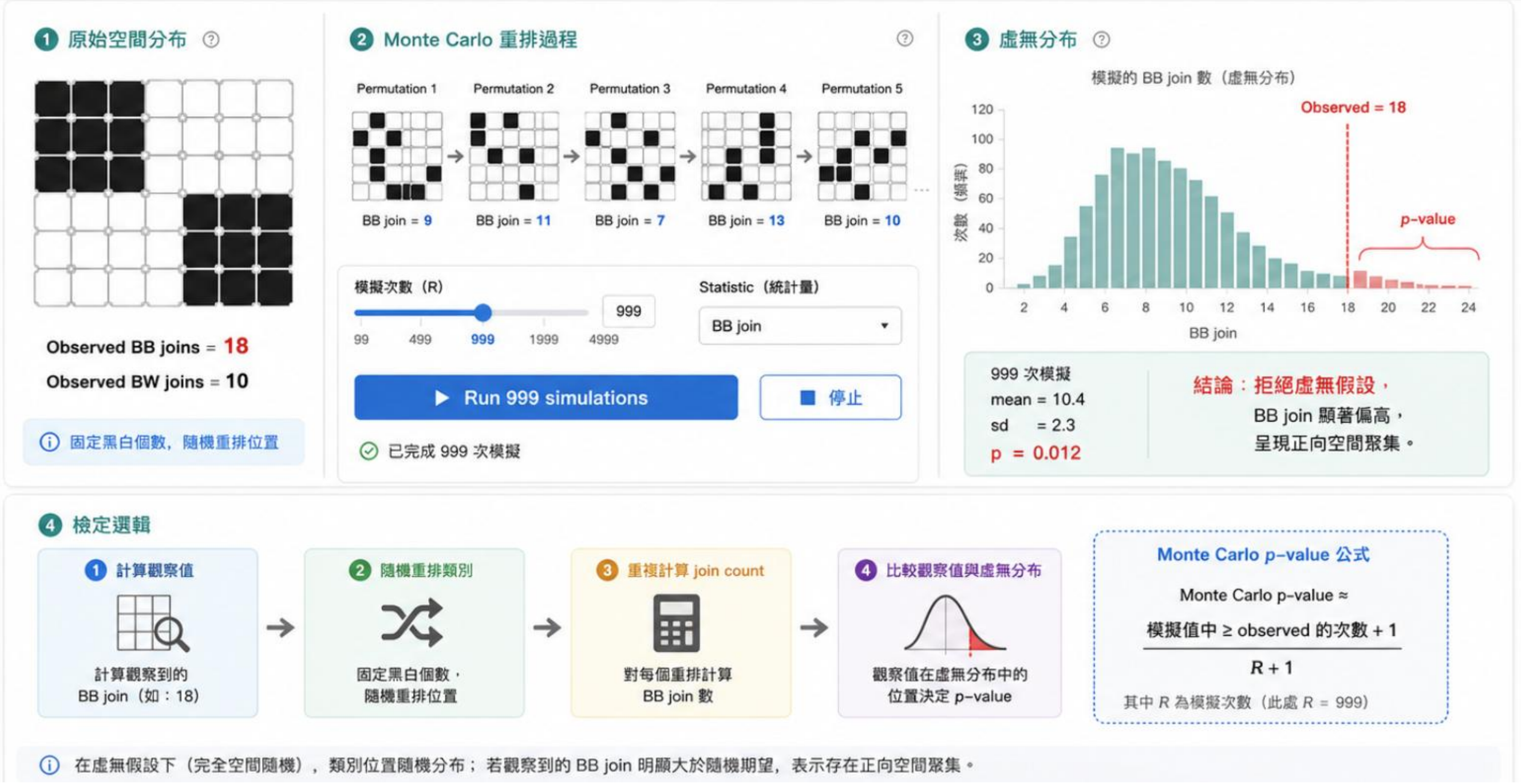
Areal Pattern A

	Areal Pattern A		
Case	Rook's	Bishop's	Queen's
BB	30	23	53
BW	8	12	20
WW	22	15	37
n	36	36	36
B	20	20	20
W	16	16	16
Total	60	50	110
E_{BB}	18.1	15.1	33.2
E_{BW}	30.5	25.4	55.9
E_{WW}	11.4	9.52	20.9
Z_{BB}	5.47	3.09	5.15
Z_{BW}	−6.02	−3.92	−7.49
Z_{WW}	5.12	2.41	4.76

Monte-Carlo Significance Test

- **Permutation test (排列檢定)**
  - The null hypothesis is that the data were determined and then assigned to their spatial locations at random.
  - The alternative is that the assignment to each location depended on the assignment at that location's neighbors.
  - The permutation test does not randomize over the possible sets of data values--it considers them given-- but **conditional on the data observed**, considers all possible ways of reassigning them to the locations.

Monte-Carlo Significance Test



① 原始空間分布

Observed BB joins = 18

Observed BW joins = 10

ⓘ 固定黑白個數，隨機重排位置

② Monte Carlo 重排過程

模擬次數（R）: 999

Statistic（統計量）: BB join ▾

✓ 已完成 999 次模擬

③ 虛無分布

999 次模擬
mean = 10.4
sd = 2.3
p = 0.012

結論：拒絕虛無假設，
BB join 顯著偏高，
呈現正向空間聚集。

④ 檢定邏輯

1. 計算觀察值：計算觀察到的 BB join（如：18）
2. 隨機重排類別：固定黑白個數，隨機重排位置
3. 重複計算 join count：對每個重排計算 BB join 數
4. 比較觀察值與虛無分布：觀察值在虛無分布中的位置決定 p-value

Monte Carlo p-value 公式

$$\text{Monte Carlo p-value} \approx \frac{\text{模擬值中} \geq \text{observed 的次數} + 1}{R + 1}$$

其中 R 為模擬次數（此處 R = 999）

ⓘ 在虛無假設下（完全空間隨機），類別位置隨機分布；若觀察到的 BB join 明顯大於隨機期望，表示存在正向空間聚集。

教學網頁： join_count_monte_carlo.html

Join Count Statistic & Monte Carlo Test

Spatial Autocorrelation · Nominal Data · Permutation-based Inference

空間統計學 互動演示

① 原理 ② Join Count ③ 虛無假說 ④ 置換邏輯 ⑤ p 值計算

● 聚集 ○ 分散 ⊞ 隨機 ▦ 棋盤

○ 清除

點擊格子切換 ■ 黑 / □ 白

觀測值 OBSERVED JOINS

BB	BW	WW
12	20	52
黑-黑	黑-白	白-白

總連結數 J = 84 | 黑格 n_B = 11 | 白格 n_W = 38

檢定設定

儉定統計量: BB（黑-黑）

寸立假設方向: 大於（聚集）

莫擬次數: 499

▶ 執行 MONTE CARLO 模擬

為什麼需要 Monte Carlo 顯著性檢定？

Join Count Statistic 是針對 **名義尺度（nominal scale）** 空間資料的自相關指標。地圖上每個單元為黑（B）或白（W），我們想知道：相鄰單元的顏色是否非隨機地聚集或分散？

理論抽樣分布（常態近似）在小樣本或極端分布時不可靠。**Monte Carlo 方法**以排列置換直接建構參考分布，不需假設，更具彈性。

Monte Carlo 參考分布

觀測值 T_OBS	模擬平均 E[T*]	模擬 SD	P 值（MC）
12	3.88	1.55	0.0020

✗ *在 $\alpha = 0.05$ 下，***拒絕 H_0***。觀測 BB = 12 顯著偏離隨機期望（偏高/聚集）（$p = 0.0020$）。空間格局具顯著的***非隨機性***。*

Case Study: Gore/Bush Presidential Election 2000

Is there evidence of clustering by State?

Total number of joins = 109

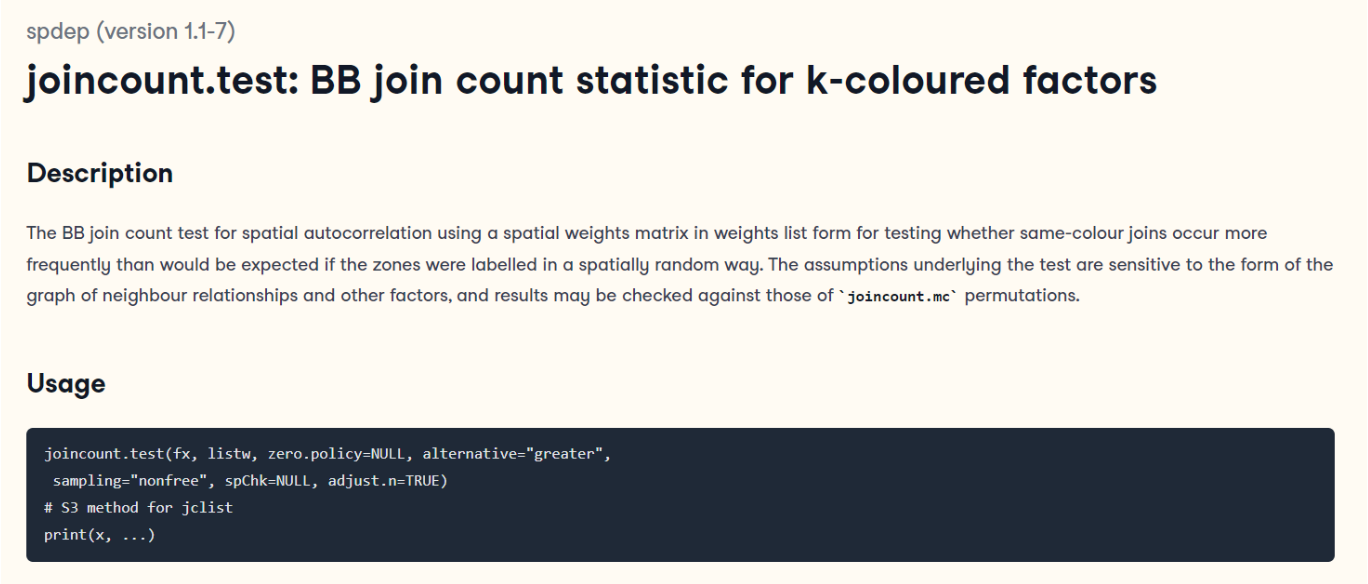
	Actual
Jbb	60
Jgg	21
Jbg	28
Total	109

Case Study: Gore/Bush Presidential Election 2000

	Actual	Expected	Stan Dev	Z-score
Jbb	60	27.125	8.667	3.7930
Jgg	21	27.375	8.704	-0.7325
Jbg	28	54.500	5.220	-5.0763
Total	109	109.000		

- There are far **more** Bush/Bush joins (actual = 60) than would be expected (27)
  - Since test score (3.79) is greater than the critical value (2.54 at 1%) result is statistically significant at the 99% confidence level (p <= 0.01)
  - Strong evidence of spatial autocorrelation—clustering
- There are far **fewer** Bush/Gore joins (actual = 28) than would be expected (54)
  - Since test score (-5.07) is greater than the critical value (2.54 at 1%) result is statistically significant at 99% confidence level (p <= 0.01)
  - strong evidence of spatial autocorrelation—clustering

R code: joincount.test() and joincount.multi()



spdep (version 1.1-7)

joincount.test: BB join count statistic for k-coloured factors

Description

The BB join count test for spatial autocorrelation using a spatial weights matrix in weights list form for testing whether same-colour joins occur more frequently than would be expected if the zones were labelled in a spatially random way. The assumptions underlying the test are sensitive to the form of the graph of neighbour relationships and other factors, and results may be checked against those of `` `joincount.mc` `` permutations.

Usage

```
joincount.test(fx, listw, zero.policy=NULL, alternative="greater",
 sampling="nonfree", spChk=NULL, adjust.n=TRUE)
# S3 method for jclist
print(x, ...)
```

spdep (version 1.1-7)

joincount.multi: BB, BW and Jtot join count statistic for k-coloured factors

Description

A function for tallying join counts between same-colour and different colour spatial objects, where neighbour relations are defined by a weights list. Given the global counts in each colour, expected counts and variances are calculated under non-free sampling, and a z-value reported. Since multiple tests are reported, no p-values are given, allowing the user to adjust the significance level applied. Jtot is the count of all different-colour joins.

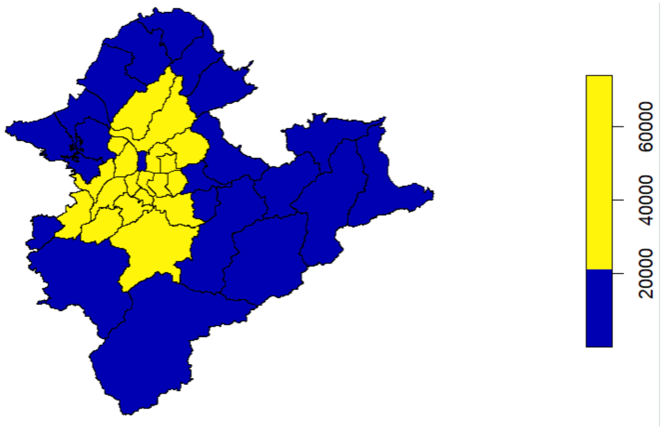
Usage

```
joincount.multi(fx, listw, zero.policy = FALSE,
 spChk = NULL, adjust.n=TRUE)
# S3 method for jcmulti
print(x, ...)
```

R code

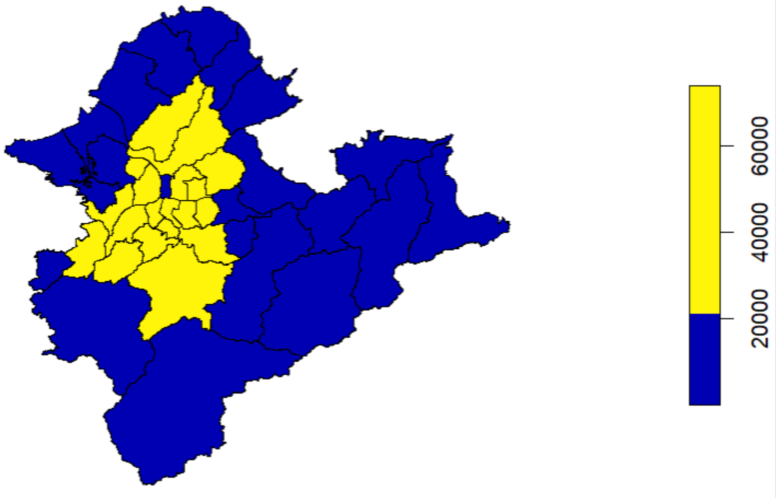
```
TWPOP_sf <- st_read("Popn_TWN2.shp",options = "encoding=Big5" )
ID <- TWPOP_sf$COUNTY_ID
Sel <- ID == "65000" | ID == "63000"
NorthTW_sf <- TWPOP_sf[Sel,]  #研究區：台北 + 新北

# 定義鄰近關係與鄰近矩陣
TWN_nb <- poly2nb(NorthTW_sf) #QUEEN = TRUE
TWN_nb_w <- nb2listw(TWN_nb, zero.policy=T)

# 計算切點：連續變數 轉換 類別變數分類
ave <- mean(NorthTW_sf$A0A14_CNT)
max <- max(NorthTW_sf$A0A14_CNT)

HIPOP <- cut(NorthTW_sf$A0A14_CNT, breaks=c(0,ave,max), labels=c("LOW","HIGH"))
names(HIPOP) <- rownames(NorthTW_sf)
joincount.test(HIPOP, TWN_nb_w)
joincount.multi(HIPOP, TWN_nb_w)
```

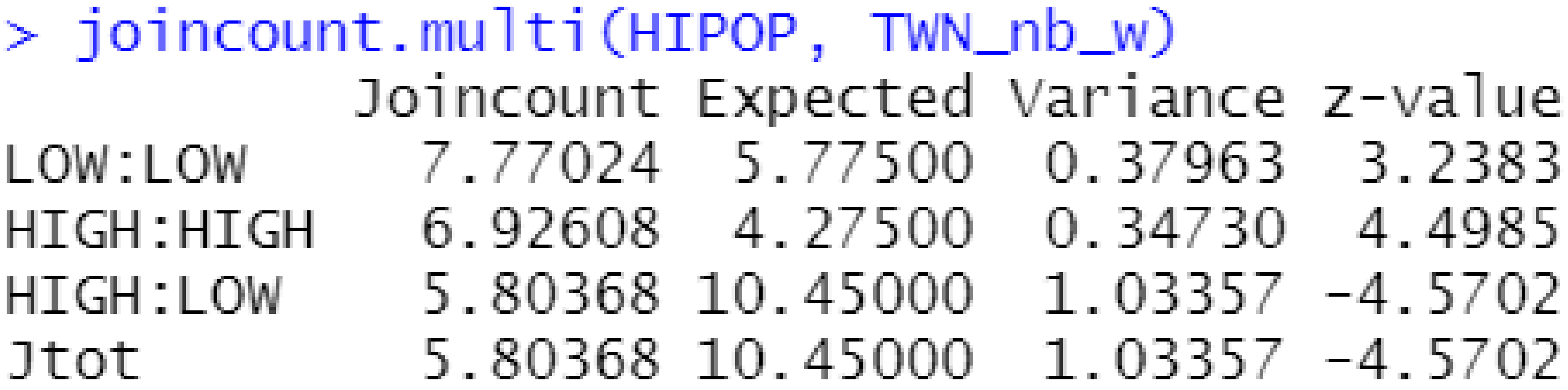
R code: Results



```
> joincount.test(HIPOP, TWN_nb_w)

        Join count test under nonfree sampling

data:  HIPOP
weights: TWN_nb_w

Std. deviate for LOW = 3.2383, p-value = 0.0006013
alternative hypothesis: greater
sample estimates:
Same colour statistic           Expectation              Variance
            7.7702381             5.7750000             0.3796304

        Join count test under nonfree sampling

data:  HIPOP
weights: TWN_nb_w

Std. deviate for HIGH = 4.4985, p-value = 3.421e-06
alternative hypothesis: greater
sample estimates:
Same colour statistic           Expectation              Variance
            6.9260823             4.2750000             0.3473002
```

R code: Results (cont'd)

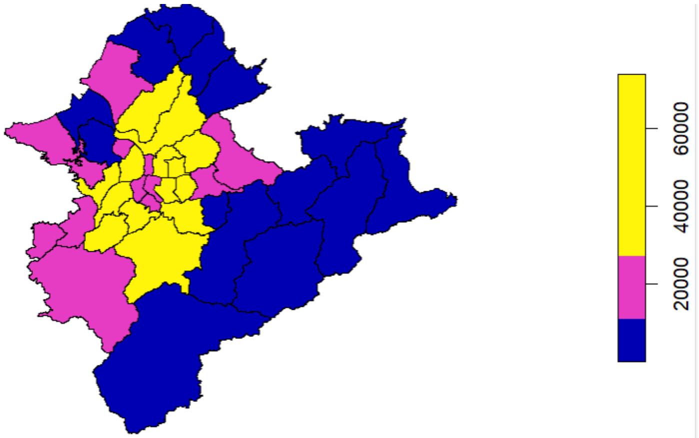




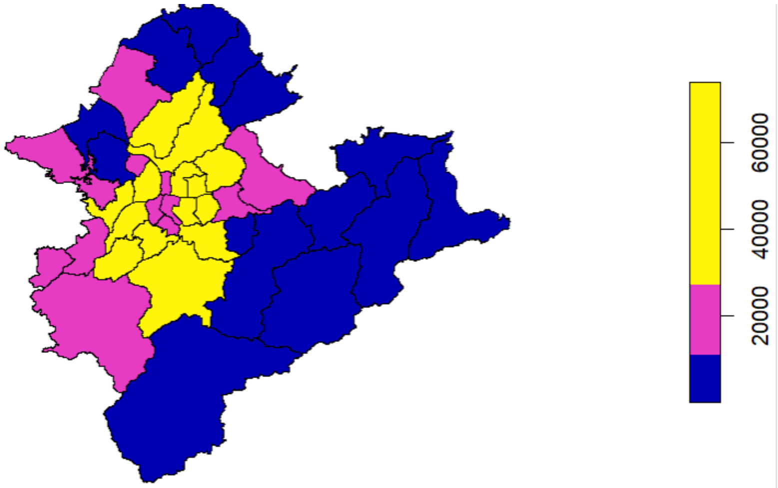
```
> joincount.multi(HIPOP, TWN_nb_w)
          Joincount Expected Variance z-value
LOW:LOW     7.77024  5.77500  0.37963  3.2383
HIGH:HIGH   6.92608  4.27500  0.34730  4.4985
HIGH:LOW    5.80368 10.45000  1.03357 -4.5702
Jtot        5.80368 10.45000  1.03357 -4.5702
```

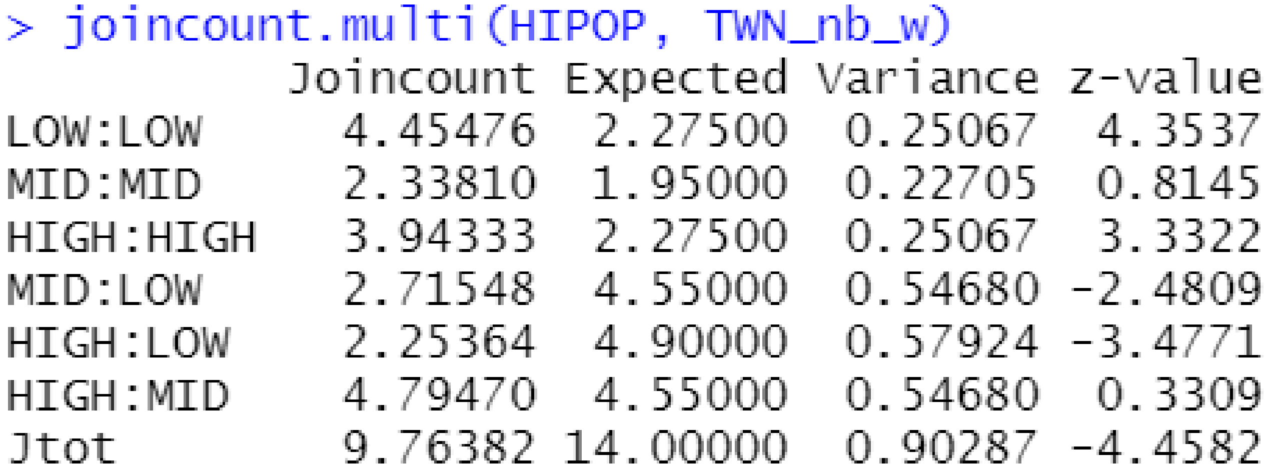
R code: 兩個以上的類別

```
cut1 <- quantile(NorthTW_sf$A0A14_CNT,0.33)
cut2 <- quantile(NorthTW_sf$A0A14_CNT,0.66)
max <- max(NorthTW_sf$A0A14_CNT)

HIPOP <- cut(NorthTW_sf$A0A14_CNT, breaks=c(0, cut1, cut2, max),
                    labels=c("LOW","MID","HIGH"))
names(HIPOP) <- rownames(NorthTW_sf)

joincount.test(HIPOP, TWN_nb_w)
joincount.multi(HIPOP, TWN_nb_w)
```

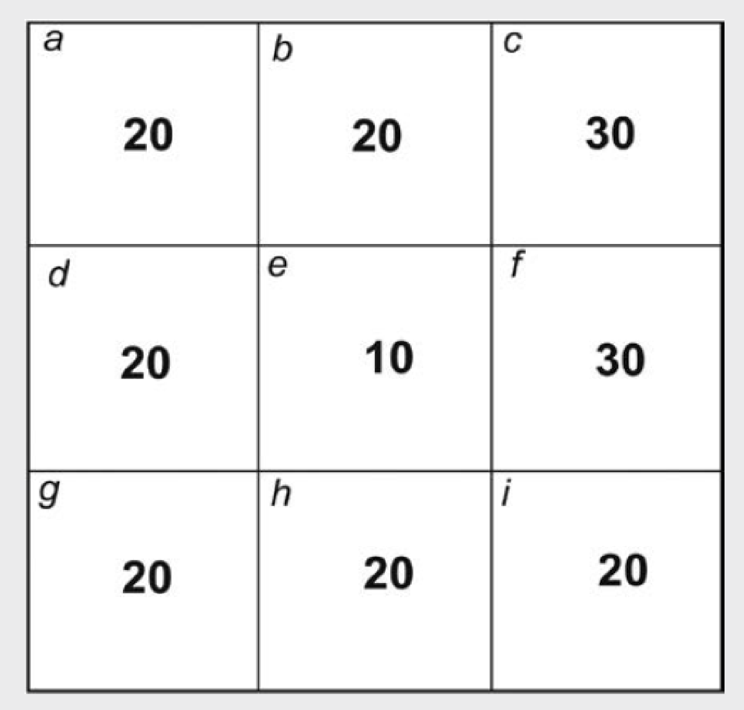


R code: Results




```
> joincount.multi(HIPOP, TWN_nb_w)
          Joincount Expected Variance z-value
LOW:LOW     4.45476  2.27500  0.25067  4.3537
MID:MID     2.33810  1.95000  0.22705  0.8145
HIGH:HIGH   3.94333  2.27500  0.25067  3.3322
MID:LOW     2.71548  4.55000  0.54680 -2.4809
HIGH:LOW    2.25364  4.90000  0.57924 -3.4771
HIGH:MID    4.79470  4.55000  0.54680  0.3309
Jtot        9.76382 14.00000  0.90287 -4.4582
```

作業 1：實作計算

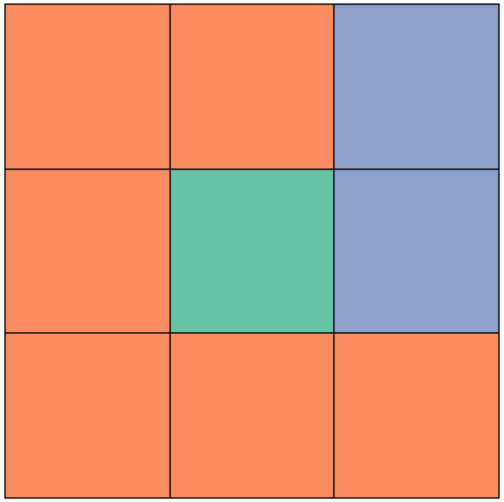


a **20**	*b* **20**	*c* **30**
d **20**	*e* **10**	*f* **30**
g **20**	*h* **20**	*i* **20**

以1st-order ROOK鄰近定義，
[1]以連續變數計算 Getis-Ord G Statistics, G(d) 以及
[2]以類別變數計算 Join Count Statistics，比較與討論檢定結果。

不能使用空間分析套件，需**利用R或Python自行撰寫程式碼**，進行統計量的計算，
並以permutation test 計算其統計顯著性。

提示：建立網格資料的sf格式

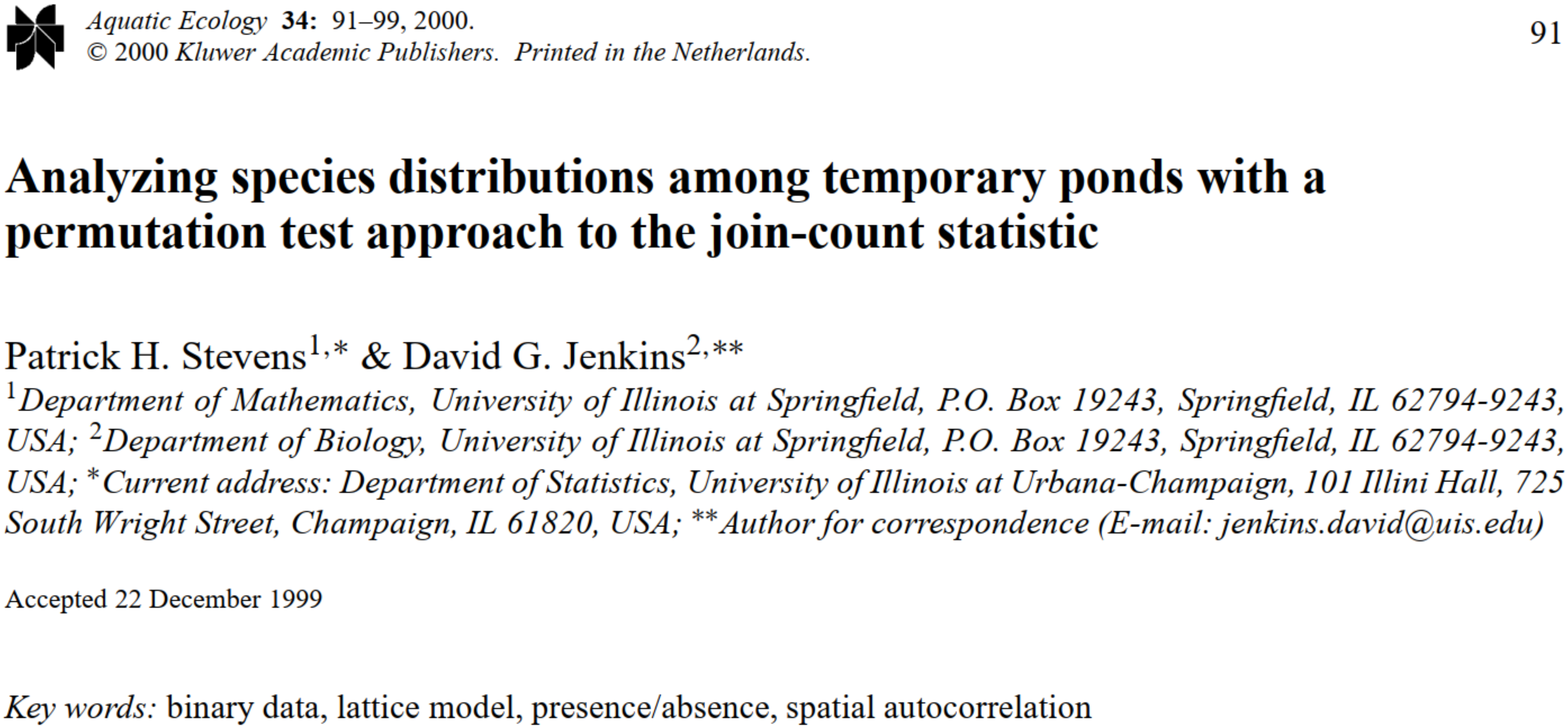


```
library(sf)

data = c(20,20,20,20,10,30,20,20,30)
sfc = st_sfc(st_polygon(list(rbind(c(0,0), c(1,0), c(1,1), c(0,0)))))
grid = st_as_sf(st_make_grid(sfc, cellsize = .4))
grid$data = data
grid$data_factor = as.factor(grid$data)

# 建立鄰近矩陣
w_list = nb2listw(poly2nb(grid,queen=F),style="B")
w = nb2mat(poly2nb(grid,queen=F),style="B")
```

作業 2：研讀論文

Reading_Join-count.Statistic.pdf

說明該篇論文如何利用join count統計方法來分析甲殼類物種的空間型態。



Aquatic Ecology **34:** 91–99, 2000.
© 2000 *Kluwer Academic Publishers. Printed in the Netherlands.*

Analyzing species distributions among temporary ponds with a permutation test approach to the join-count statistic

Patrick H. Stevens[1,*] & David G. Jenkins[2,**]
[1]*Department of Mathematics, University of Illinois at Springfield, P.O. Box 19243, Springfield, IL 62794-9243, USA;* [2]*Department of Biology, University of Illinois at Springfield, P.O. Box 19243, Springfield, IL 62794-9243, USA;* ***Current address: Department of Statistics, University of Illinois at Urbana-Champaign, 101 Illini Hall, 725 South Wright Street, Champaign, IL 61820, USA;* ***Author for correspondence (E-mail: jenkins.david@uis.edu)*

Accepted 22 December 1999

Key words: binary data, lattice model, presence/absence, spatial autocorrelation

統計圖表解讀

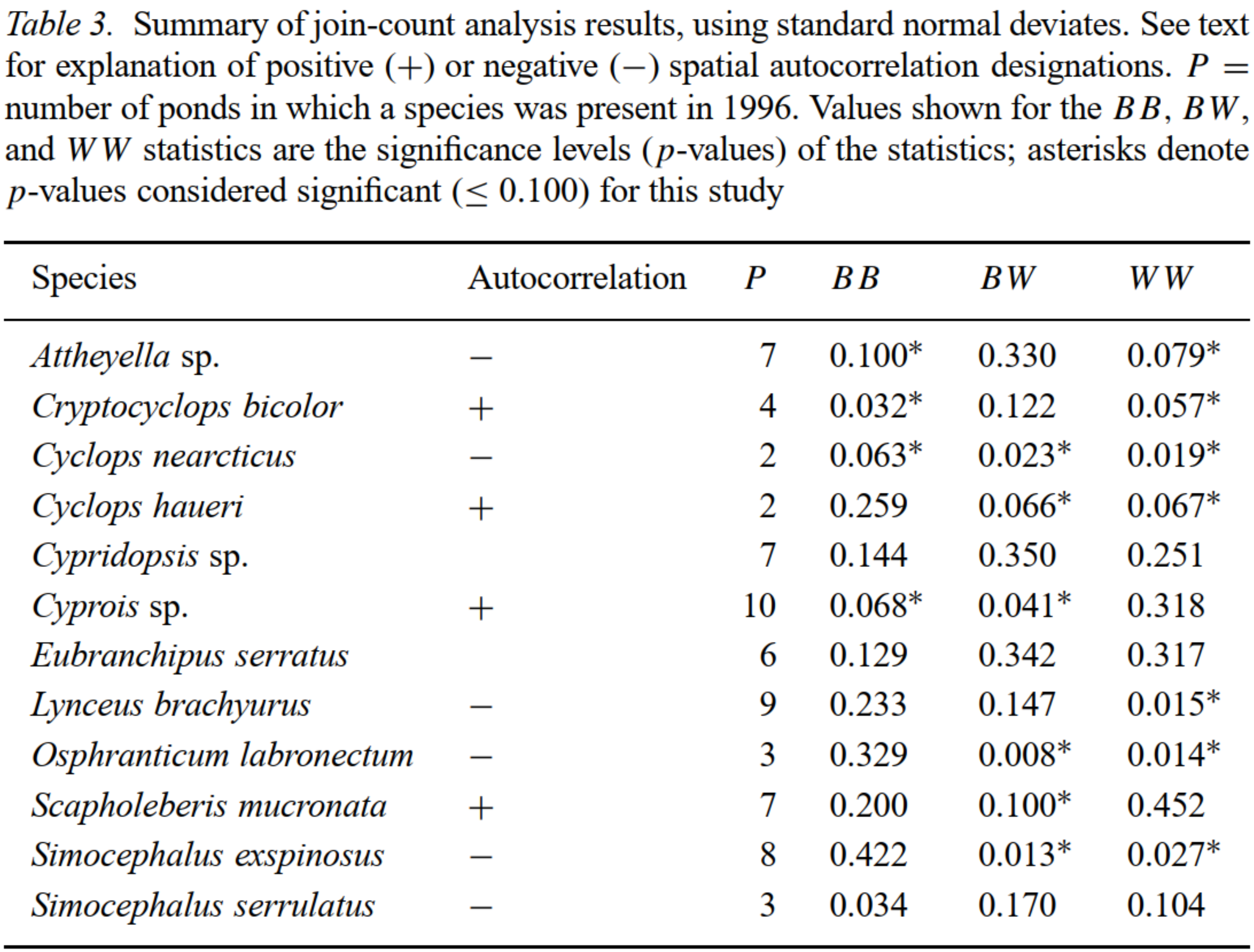


Table 3. Summary of join-count analysis results, using standard normal deviates. See text for explanation of positive (+) or negative (−) spatial autocorrelation designations. P = number of ponds in which a species was present in 1996. Values shown for the BB, BW, and WW statistics are the significance levels (p-values) of the statistics; asterisks denote p-values considered significant (≤ 0.100) for this study

Species	Autocorrelation	P	BB	BW	WW
Attheyella sp.	−	7	0.100*	0.330	0.079*
Cryptocyclops bicolor	+	4	0.032*	0.122	0.057*
Cyclops nearcticus	−	2	0.063*	0.023*	0.019*
Cyclops haueri	+	2	0.259	0.066*	0.067*
Cypridopsis sp.		7	0.144	0.350	0.251
Cyprois sp.	+	10	0.068*	0.041*	0.318
Eubranchipus serratus		6	0.129	0.342	0.317
Lynceus brachyurus	−	9	0.233	0.147	0.015*
Osphranticum labronectum	−	3	0.329	0.008*	0.014*
Scapholeberis mucronata	+	7	0.200	0.100*	0.452
Simocephalus exspinosus	−	8	0.422	0.013*	0.027*
Simocephalus serrulatus	−	3	0.034	0.170	0.104